

**REMARKS**

Applicant respectfully requests reconsideration and allowance of the subject application. Claims 1, 4-57 are pending in this application.

**Claim Amendments**

Claims 1-57 were previously pending.

Claims 1, 4, 6, 10, 49, and 54 are currently amended.

No new claims are added.

Claims 2-3 are cancelled.

Claims 1, 4-57 are pending.

**Rejection of the Claims****Rejections under 35 USC § 102(b)**

Claims 1-34 and 40-57 were rejected under 35 USC § 102(b) as being anticipated by U.S. Patent No. 5,572,423 to Church, filed January 23, 1995 ("Church" or "the Church reference").

**Claim 1**

Although Applicant respectfully disagrees with the Examiner's characterization of the Church reference and characterization of claim 1, Applicant nonetheless amends claim 1 to more particularly point out and distinctly claim the subject matter. The current amendment does not narrow claim 1 but merely adds clarity.

Claim 1, as amended, defines a method for determining the likelihood that a word in a dictionary is being incorrectly represented by a string; comprising:

- iteratively partitioning the word into multiple segments, each segment consisting of a character or character sequence, wherein each iteration partitions the word into a different number of the multiple segments;

- for each iteration of the partitioning, iteratively varying the lengths of the segments while maintaining the number of the segments;

- for each iteration of the partitioning, dividing the string into the same number of string segments as the number of word segments and iteratively varying the lengths of the string segments, wherein corresponding word segments and string segments can be of different lengths;

- for each iteration of varying the lengths of the word segments and the string segments, computing a probability for each pair, wherein each pair consists of a word segment and a corresponding string segment, and wherein the probability consists of a likelihood that the word segment is being incorrectly represented by the string segment;

- for each iteration of varying the lengths, computing a product of the probabilities of the pairs; and

- determining the likelihood that the word is being incorrectly represented by the string based on one of the products.

Applicant's specification states:

"the error model 44 partitions the word *w* and string *s* into different numbers of segments that define varying lengths of character sequences. For example, suppose the dictionary word is "physical" and the number of partition segments is five. One possible partition is, say, "ph y s ic al". (Applicants' specification, page 9, lines 16-19; emphasis added.)

Thus, in one implementation, Applicant's subject matter partitions an entered string and a candidate word, each into the same number of segments. During iterations, the method partitions the word and the string into different numbers of segments, but for any iteration the entered string and the candidate word are each divided into the same number of segments.

Corresponding segments in the word and the string can be the same or different lengths. In fact, while the word and the string are divided into a given number of segments, the method tries various combinations of segment lengths in both the word and the string. So, for example, in one iteration of partitioning, the word and string may be partitioned into five segments, but various combinations of segment lengths are tried while the partitioning still consists of five segments. After considering the various segment length combinations possible under a partition of five segments, the method may move on to a partition of six segments, etc.

Because the method calculates probabilities of a myriad of pairs, each consisting of a word segment and a corresponding string segment, and because the segments in a pair can each be character sequences of different lengths, Applicant's method can calculate certain probabilities that describe converting the word into the string or vice versa, that are not possible in the conventional references cited by the Office. Thus, if the word is "schnitzel" and the string is "snitzel" the string segment "s" may be variously paired with the word segments "s", "sc", "sch" during the different iterations of partitioning and segment varying.

During a given iteration of varying the length of segments, the probability that each set of word segments is being incorrectly represented by a corresponding set of string segments is calculated. When all the iterations of partitioning and all the iterations of varying the lengths of the segments are performed, then the highest probability(ies) obtained can be compared with probabilities of other candidate words that the misspelled string might be intending to represent, to rank these candidate words, or to select a best candidate word. In other words, the process of partitioning the string and the word different ways and weighing the

relative merit of each way is typically repeated many times. (See Table 1 on page 11 of Applicants' specification.) The partitioning with the highest merit can be selected for comparison with the best partitioning merits of other candidate words.

### The Church reference

The Church reference, on the other hand, does not expressly or inherently describe the subject matter of claim 1. Applicant respectfully points out that pages 3-4 of Applicant's specification discuss the Church model as prior art at some length. For example:

The error model probability  $P(s|w)$  used in noisy channel spell correction programs, such as the one described in Church et al., may seem backwards initially because it suggests finding how likely a string  $s$  is to be entered given that a dictionary word  $w$  is intended. In contrast, the spell correction program actually wants to know how likely the entered string  $s$  is to be a word  $w$  in the dictionary, or  $P(w|s)$ . The error model probability  $P(s|w)$  comes from Bayes formula, which can be used to represent the desired probability  $P(w|s)$  as follows:"

$$P(w|s) = \frac{P(s|w) \cdot P(w)}{P(s)}$$

The denominator  $P(s)$  remains the same for purposes of comparing possible intended words given the entered string. Accordingly, the spell checking analysis concerns only the numerator product  $P(s|w) \cdot P(w)$ , where the probability  $P(s|w)$  represents the error model and the probability  $P(w)$  represents the source model.

Whereas Church presents a conventional noisy channel spelling correction scheme, Applicant's subject matter provides a greatly improved error model (that is, the " $P(s|w)$ " term in the numerator of the Bayes formula above).

The Office presents Fig. 1 of Church in support of the Office's 35 USC § 102(b) rejection. However, Church merely uses "the identical set of edit types

used by Mays et al. (i.e., single letter insertion, substitution, deletion, and letter-pair transposition)..." see Applicant's specification, page 3. The Examiner's cited conversion of "acress" to "actress" is a single character insertion of a "t". If instead of "acress" the string "ackterris" is the misspelled form of "actress," the Church scheme would not function, because the Church scheme has no mechanism for obtaining a probability that describes the conversion of multi-character sequences that need to be substituted in order to obtain the intended word (in this example, notably "ck" to "c" and "errr" to "r").

Thus, the Church reference does not disclose many of the elements of claim 1, such as iteratively partitioning the word into multiple segments; iteratively varying the lengths of the segments in the word and the string while maintaining the number of segments; calculating probabilities of various combinations of word segment/string segment pairs (where a word segment and a string segment of the pair might be multiple character sequences of different lengths from each other) in order to find the best pairs yielding the most likely word, etc.

Applicant therefore respectfully suggests the 35 USC § 102(b) rejection based on the Church reference be removed, and that claim 1 is in condition for allowance.

### Claim 2-3

Applicant requests that claims 2 and 3 be canceled without prejudice. The subject matter of claims 2-3 has been moved directly or indirectly to claim 1, to add clarity to claim 1.

Claims 4-9

For at least the reasons set forth above with respect to claim 1, Applicant submits that claims 4-9 are patentable over the Church reference. Dependent claims contain the language of the claims from which they depend. Claims 4-9 depend from claim 1. Therefore, claims 4-9 are also allowable.

Claim 10

Claim 10 has been amended to more particularly point out and distinctly claim the subject matter.

Independent claim 10, as amended, recites a method comprising:

receiving an entered string  $s$ ; and  
determining a probability  $P(s|w)$  expressing how likely a word  $w$  was to have been incorrectly entered as the string  $s$  based on partitioning the word  $w$  and the string  $s$  and computing probabilities for various partitionings, wherein a probability for a partitioning represents the probability that one or more edit operations convert first arbitrary-length character sequences  $\alpha_1, \alpha_2, \alpha_3, \dots, \alpha_n$  in the word  $w$  to corresponding second arbitrary-length character sequences  $\beta_1, \beta_2, \beta_3, \dots, \beta_n$  in the string  $s$ , wherein:

$$P(s|w) = P(\beta_1|\alpha_1) * P(\beta_2|\alpha_2) * P(\beta_3|\alpha_3) * \dots * P(\beta_n|\alpha_n) .$$

The Church reference, on the other hand, does not expressly or inherently disclose determining a probability  $P(s|w)$  expressing how likely a word  $w$  was to have been incorrectly entered as the string  $s$  based on partitioning the word  $w$  and the string  $s$  and computing probabilities for various partitionings, etc. Therefore, Applicant respectfully requests that the rejection under 35 USC § 102(b) be removed, and that claim 10 is in condition for allowance.

Claims 11-16

For at least the reasons set forth above with respect to claim 10, Applicant submits that claims 11-16 are patentable over the Church reference. Dependent claims contain the language of the claims from which they depend. Claims 11-16 depend from claim 10. Therefore, claims 11-16 are also allowable.

Claim 17

Claim 17 defines a method of:

receiving an entered string  $s$ ; and  
determining a probability  $P(s|w)$  expressing how likely a word  $w$  was to have been incorrectly entered as the string  $s$ , by partitioning the word  $w$  and the string  $s$  and computing probabilities for various partitionings, as follows:

$$P(s | w) = \sum_{R \in \text{Part}(w)} P(R | w) \sum_{\substack{T \in \text{Part}(s) \\ |T|=|R|}} \prod_{i=1}^{|R|} P(T_i | R_i)$$

where  $\text{Part}(w)$  is a set of possible ways of partitioning the word  $w$ ,  $\text{Part}(s)$  is a set of possible ways of partitioning the string  $s$ ,  $R$  is a particular partition of the word  $w$ , and  $T$  is a particular partition of the string  $s$ .

The Church reference, on the other hand, does not expressly or inherently disclose partitioning of the probability function in the misspelled word and candidate word. Thus, Applicant respectfully requests that the rejection of claim 17 under 35 USC § 102(b) as being anticipated by Church, be removed.

The Office states:

“Church as applied to claims 1-9 and 10-16 above, discloses every aspect of Applicant’s claimed invention except for not explicitly disclosing

partitioning of the probability functions in the misspelled word and candidate word. However Church discloses the computing probabilities  $P(\text{acress}|\text{actress})$  which is identical to the present invention  $P(\text{string}|\text{word})$ . Therefore the computing probabilities' outcome of Church and the claimed invention are the same."

The Office points out above that Church does not explicitly disclose partitioning of the probability functions in the misspelled word and candidate word. Claim 17, on the contrary, defines a method that partitions the probability functions in the misspelled word and candidate word. Thus, Applicant respectfully requests that the rejection of claim 17 under 35 USC § 102(b) as being anticipated by Church, be removed.

Also, since Church does not disclose partitioning of the probability functions in the misspelled word and candidate word then it does not follow, as the Office asserts, that the computing probabilities  $P(\text{acress}|\text{actress})$  is identical to the present invention  $P(\text{string}|\text{word})$ . Rather, the opposite is true. Since Church does not disclose partitioning of the probability functions in the misspelled word and candidate word "the computing probabilities' outcome of Church and the claimed invention" are probably not the same, and even further, probably cannot be the same.

Thus, Applicant again respectfully requests that the rejection of claim 17 under 35 USC § 102(b) be removed for the reasons just given and for the reasons given above for claims 1 and 10.



Claims 18-22

For at least the reasons set forth above with respect to claim 17, Applicant submits that claims 18-22 are patentable over the Church reference. Dependent claims contain the language of the claims from which they depend. Claims 18-22 depend from claim 17. Therefore, claims 18-22 are also allowable.

Claim 23

Claim 23 was rejected by the Office under the same rationale as given by the Office for rejecting claims 17-22. Thus, Applicant respectfully requests that the rejection of claim 23 under 35 USC § 102(b) be removed for the reasons given above for claims 1, 10, and 17.

Claims 24-29

For at least the reasons set forth above with respect to claim 23, Applicant submits that claims 24-29 are patentable over the Church reference. Dependent claims contain the language of the claims from which they depend. Claims 24-29 depend from claim 23. Therefore, claims 24-29 are also allowable.

Claim 30

Claim 30 was rejected by the Office under the same rationale as given by the Office for rejecting claims 17-22. Thus, Applicant respectfully requests that the rejection of claim 30 under 35 USC § 102(b) be removed for the reasons given above for claims 1, 10, and 17.

Claims 31-34

For at least the reasons set forth above with respect to claim 30, Applicant submits that claims 31-34 are patentable over the Church reference. Dependent claims contain the language of the claims from which they depend. Claims 31-34 depend from claim 30. Therefore, claims 31-34 are also allowable.

Claim 40

Claim 40 was rejected by the Office under the same rationale as given by the Office for rejecting claim 1. Thus, Applicant respectfully requests that the rejection of claim 40 under 35 USC § 102(b) be removed for the reasons given above for claim 1.

Claims 41-48

For at least the reasons set forth above with respect to claim 40, Applicant submits that claims 41-48 are patentable over the Church reference. Dependent claims contain the language of the claims from which they depend. Claims 41-48 depend from claim 40. Therefore, claims 41-48 are also allowable.

Claim 49

Claim 49 has been amended to more particularly point out and distinctly claim the subject matter. The amendment does not narrow the claim but rather adds clarity. Claim 49 defines a program embodied on a computer readable medium, which when executed, directs a computer to perform the following:

- (a) receive an entered string  $s$ ;
- (b) for multiple words  $w$  in a dictionary, determine:
  - how likely a word  $w$  in a dictionary is to have been generated,  $P(w|context)$ ; and
  - how likely the word  $w$  was to have been entered as the string  $s$ ,  $P(s|w)$ , based on partitioning the word  $w$  and the string  $s$  and computing probabilities for various partitionings to determine a highest likelihood of at least one edit operation that converts one of multiple character sequences of arbitrary length in the word to one of multiple character sequences of arbitrary length in the string; and
- (c) maximize  $P(s|w)*P(w|context)$  to identify which of the words is most likely the word intended when the string  $s$  was entered.

The Church reference does not expressly or inherently describe the subject matter of claim 49, for example, does not describe determining how likely a word was to have been entered as a string based on partitioning the word  $w$  and the string  $s$  and computing probabilities for various partitionings.

Since the Church reference does not expressly or inherently describe the entire subject matter of claim 49, Applicants respectfully request that the 35 USC § 102(b) rejection be removed. Applicants further suggest that claim 49 is in condition for allowance.

#### Claims 50-53

For at least the reasons set forth above with respect to claim 49, Applicant submits that claims 50-53 are patentable over the Church reference. Dependent claims contain the language of the claims from which they depend. Claims 50-53 depend from claim 49. Therefore, claims 50-53 are also allowable.

#### Claim 54

Claim 54 has been amended to more particularly point out and distinctly claim the subject matter. Claim 54 defines a spell checker that includes:

a source model component to determine how likely a word  $w$  in a dictionary is to have been generated; and

an error model component to determine how likely the word  $w$  was to have been incorrectly entered as the string  $s$  based on arbitrary length string-to-string transformations, wherein the error model partitions the word  $w$  and the string  $s$  and computes probabilities for various partitionings.

The Church reference does not expressly or inherently describe the subject matter of claim 54, for example, does not describe determining how likely a word was to have been entered as a string based on partitioning the word  $w$  and the string  $s$  and computing probabilities for various partitionings.

Since the Church reference does not expressly or inherently describe the entire subject matter of claim 54, Applicants respectfully request that the 35 USC § 102(b) rejection be removed. Applicants further suggest that claim 54 is in condition for allowance.

#### Claims 55-57

For at least the reasons set forth above with respect to claim 54, Applicant submits that claims 55-57 are patentable over the Church reference. Dependent claims contain the language of the claims from which they depend. Claims 55-57 depend from claim 54. Therefore, claims 55-57 are also allowable.

**Rejections under 35 USC § 103(a)**

Claims 35-39 were rejected as being unpatentable over U.S. Patent No. 6,131,102 to Potter (“Potter” or “the Potter reference”) in view of the Church reference.

Applicants’ claim 35 recites:

A method for training an error model used in a spell checker, comprising:

determining, given a <wrong, right> training pair and multiple single character edits that convert characters in one of the right or wrong strings to characters in the other of the right or wrong strings at differing costs, an alignment of the wrong string and the right string that results in a least cost to convert the characters;

collapsing any contiguous non-match edits into one or more common error regions, each error region containing one or more characters that can be converted to one or more other characters using a substitution edit; and  
computing a probability for each substitution edit.

Applicant notes that the Office characterizes Potter as using “multiple single character edits” and the Office states that Potter does not explicitly disclose “collapsing any contiguous non-match edits into one or more common error regions, each error region containing one or more characters that can be converted to one or more other characters using a substitution edit; and computing a probability for each substitution edit.”

Moreover, the Patent Office characterizes Table D of the Church reference as disclosing “computing a probability for each substitution edit.” But Church does not disclose, teach or suggest alone or in combination with the Potter reference, the subject matter of claim 35, in particular, “collapsing any contiguous non-match edits into one or more common error regions....” Thus, neither Potter nor Church singly or in combination suggest a method for training an error model

to use substitution of entire character sequences of arbitrary length within strings and words, e.g., “ct” for “kgs”; “act” for “akgs”; “actu” for “akgsu”; “ctu” for “kgsu”; “ctua” for “kgsua”; etc. (Applicants’ specification, pages 16-17.)

Assuming for the sake of argument that a motivation to combine the references existed, the Potter and Church references do not suggest a training mechanism that can use character sequences of arbitrary length for a single substitution because both Potter and Church rely on single character edits. The Office cites examples “D” and “E” of Church’s table D as reading on Applicant’s “collapsing any contiguous non-match edits into one or more common error regions, each error region containing one or more characters that can be converted to one or more other characters using a substitution edit; and computing a probability for each substitution edit.” However, example D in Church’s table changes a misspelled string (“acress”) into a first candidate word (“access”) while example E changes the misspelled string (“acress”) into a second candidate word (“across”). Thus, Church is talking about two single character edits leading to two different words, not two single character edits somehow combined to lead to a single word. In other words, Church’s examples D and E cannot be combined to somehow suggest Applicant’s substitution of entire character sequences of arbitrary length within strings and words.

But the Patent Office also presents no evidence of motivation to combine Potter and Church to arrive at Applicants’ subject matter. There would be no motivation to combine Potter’s single character edits with Church’s probabilities for single character edits to arrive at Applicants’ claim 35.

The Patent Office has not established a prima facie case of obviousness for claim 35 because Church does not disclose, teach or suggest alone or in combination with the Potter reference, the subject matter of claim 35, in particular

the collapsing any contiguous non-match edits into one or more common error regions, each error region containing one or more characters that can be converted to one or more other characters using a substitution edit (emphasis added).

Applicant therefore respectfully submits that the 35 USC § 103(a) rejection is traversed and that claim 35 is patentable and nonobvious over Potter in view of Church and is therefore in condition for allowance.

#### Claims 36-39

For at least the reasons set forth above with respect to claim 35, Applicant submits that claims 36-39 are also nonobvious and patentable over the Potter reference in view of the Church reference. Dependent claims contain the language of the claims from which they depend. Claims 36-39 depend from claim 35. Therefore, claims 36-39 are also allowable.

**CONCLUSION**

Applicant respectfully suggests that claims 1, 4-57 are in condition for allowance. Applicant respectfully requests reconsideration and issuance of the subject application. Should any matter in this case remain unresolved, the undersigned attorney respectfully requests a telephone conference with the Examiner to resolve any such outstanding matter.

Respectfully Submitted,

Date: 4-4-05

By: Mark C. Farrell  
Lee & Hayes PLLC  
Mark C. Farrell  
Reg. No. 45,988  
(509) 324-9256